Project	**IEEE 802.16 Broadband Wireless Access Working Group**	
Title	**An Efficient Media Access Control Protocol for Broadband Wireless Access Systems**	
Date Submitted	**1999-10-28**	
Source	Dr. James F. Mollenauer Technical Strategy Associates 37 Silver Birch Road Newton, Massachusetts, 02468	Voice: 617-244-0077 Fax: 617-244-0077 E-mail: jmollenauer@TechnicalStrategy.com
	Jay Klein Ensemble Communications, Inc. 6256 Greenwich Dr., Ste 400 San Diego, CA 92122	Voice: 858-404-6544 Fax: 858-458-1401 E-mail: jay@ensemblecom.com
	Brian Petry 3Com Corp. 12230 World Trade Dr. San Diego, CA 92128	Voice: 858-675-8533 E-mail: brian_petry@3com.com
Re:	This is a MAC protocol proposal for "Phase I" of the 802.16 MAC protocol selection process. It is in response to the call for contributions, 802.16p-99/1, as input to 802.16 Session #4.	
Abstract	The proposal described herein describes a MAC protocol that: • Supports the transport of diverse traffic types simultaneously (TDM, variable- and fixed-length PDU) • Maximizes capacity of the air link • Provides a commercially viable network for system operators • Uses well-understood technology • Supports FDD (both full and half-duplex) and TDD • Is responsive to varying bandwidth demands The MAC protocol closely resembles cable modem MAC protocols (e.g., DOCSIS and 802.14) yet addresses issues important to BWA systems.	
Purpose	The 802.16 Working Group should consider this MAC protocol proposal at Session #4.	
Notice	This document has been prepared to assist the IEEE 802.16. It is offered as a basis for discussion and is not binding on the contributing individual(s) or organization(s). The material in this document is subject to change in form and content after further study. The contributor(s) reserve(s) the right to add, amend or withdraw material contained herein.	
Release	The contributor acknowledges and accepts that this contribution may be made public by 802.16.	
IEEE Patent Policy	The contributor is familiar with the IEEE Patent Policy, which is set forth in the IEEE-SA Standards Board Bylaws <http://standards.ieee.org/guides/bylaws> and includes the statement: "IEEE standards may include the known use of patent(s), including patent applications, if there is technical justification in the opinion of the standards-developing committee and provided the IEEE receives assurance from the patent holder that it will license applicants under reasonable terms and conditions for the purpose of implementing the standard."	

An Efficient Media Access Control Protocol for Broadband Wireless Access Systems

Dr. James F. Mollenauer
Jay Klein
Brian Petry

1 Introduction: Goals for the MAC Protocol

The MAC for Broadband Wireless Access must meet a number of goals simultaneously:

- Support diverse traffic types simultaneously:
  - TDM, variable length, fixed length
- Maximize capacity of the air link
- Provide a commercially viable network for system operators
- Use well-understood technology
- Support FDD (both full and half-duplex) and TDD
- Be responsive to varying bandwidth demands

We will discuss these goals individually and then describe a protocol which meets the goals in an optimal way.

1.1 Support diverse traffic types simultaneously

Wireless access networks provide a service equivalent to that of fiber in linking the customer's premises to backbone networks and service providers. Since customers have a variety of traffic types flowing through their existing links, it follows that a wireless access provider should be able to duplicate these facilities, but with the cost savings that accompany the use of wireless rather than copper or glass. Any significant data type that is not carried over the wireless link, or which requires expensive conversion, reduces the value of the 802.16 system.

The traffic types include:

1.1.1 TDM

This is the predominant format today for access networks. It includes familiar trunk lines like T1 and E1, which were designed for voice but are often used for data applications as well as for fax and video conferencing. This usage is unpredictable: any line (or DS0 channel within T1) may be used sometimes with a modem or possibly a fax machine. Therefore it is not possible to apply a strategy of treating the traffic as voice and compressing it, as might be done for voice over IP. It is important, moreover, to preserve the low-delay characteristics of hard-wired access links.

While this type of traffic may be on the decline, relative to others, it will remain very large in quantity for some time. It is important for service providers to be able to take over existing access lines as part of their service offerings, without waiting until their potential customers migrate to some other form of access.

1.1.2 Variable-length packets

Packets are the fastest-growing segment of the traffic, particularly IP packets. There are also other packet formats in common use: Frame Relay, IPX, and SNA to name a few. All of these should be carried over the air link with as little conversion overhead as possible. Despite the recent surge of interest in voice over IP, it is unlikely that this will become the predominant form of traffic on corporate access networks in less than a decade.

1.1.3 Fixed-length data units

The predominant format in this category, of course, is ATM cells. ATM is important to access networks, not so much because of its usage on customer premises, but because of its importance in carrier backbones. Even IP traffic is often converted to ATM by carriers for travel over their backbones.

ATM provides well for a variety of service types. ATM networks usually are capable of supplying a variety of quality-of-service levels: constant bit rate, variable bit rate (both real-time and non-real-time), available bit rate, and unspecified bit rate. It is important that the access network be able to provide equivalent class-of-service differentiation, so that the BWA air link can be one link in an ATM network in which the end points are not part of the 802.16 network.

1.2 Maximize capacity of the air link

Bandwidth over the air link is a scarce commodity, or will be soon, if experience with all other media is duplicated. Therefore a great deal of attention must be paid to efficiency. To this end, this proposal involves several methods that maximize efficiency:

1.2.1 Compression

Several compression mechanisms are involved. One of these is to compress TDM channels carried over the air link. Typically T1 and E1 trunks are not fully utilized in order not to deny outgoing and incoming calls an available line. Compression involves intercepting the signaling which takes place between the PBX and the central office so as to know which channels are in use (off-hook) and which are unused. The unused ones are not carried, leaving capacity which may be assigned on a demand basis to other applications.

Similarly, ATM links pay a “cell tax” of about 10% because of cell headers. In this proposal, ATM headers are compressed by sending only the first header of a set of cells with the same header. The end-of-packet bit in AAL5 is encoded in the run-length field.

1.2.2 Multiple modulation types

Because the radio signal drops in strength with distance, it is possible to transmit to nearby users with higher-order constellations than to fringe users, at a given error ratio. This makes it possible to reduce the air time consumed by the nearby users. Typically, a mix of users working with QAM-4, -16, and –64 modulations can be supported at the same time. Should propagation conditions change, the modulation type for any user can be changed accordingly.

1.2.3 Support Various Duplex Options

At the MAC level the differences between TDD and half- or full-duplex FDD are minor, affecting mainly the time association between downstream and upstream. As various duplex schemes have their own merits coupled with the fact that the target spectrum options worldwide are enormous, the MAC should be designed to address all duplex options.

1.2.4 Continuing allocations and piggybacking

The overhead of contention in the making of requests is reduced by using a continuing-allocation strategy for constant-rate applications and by allowing piggybacking of new requests during already-allocated transmission times.

1.3 Provide a commercially-viable network for system operators

Several features are incorporated in this proposal which are necessary in running a third-party public service. This aspect of the proposal is very similar to the case with cable modems, in which the cable company likewise operates a service for paying customers.

1.3.1 Authentication

This is needed to establish the legitimacy of the user, and to provide the resource-allocation components of the network with information as to the contractual service level for the particular customer. Such policies can are established by the service operator, but enforcement of the service level agreement (SLA) is the province of the MAC.

1.3.2 Encryption

With user passwords and data going through the air and accessible to anyone with a suitable antenna, privacy is a must. Encryption is used on all user data and during appropriate portions of the registration process. As with cable modems, encryption is based on use of public-key technology for authentication and key distribution, in combination with DES or triple DES for the data itself.

1.3.3 Centralized resource allocation

Public networks are not peer networks. All users do not necessarily have the same transmission rights, and means must be provided within the MAC to enforce the SLA and to deny access to those whose actions could impact the service provided to other users.

This centralized allocation of bandwidth also allows other types of networks (ATM or TCP/IP with RSVP, etc.) to operate over the 802.16 network.

1.4 Use well-understood technology

The technology described in this proposal has been implemented and is currently under test operation.

Some aspects of the proposal are taken from cable modem practice, but it should be emphasized that there are important differences between cable modem networks and BWA systems. These include the fact that BWA is targeted to a much larger degree on commercial customers rather than homes. One consequence of this is that TDM traffic is important for BWA, while it is not for home users. Another example is that cable systems utilize fiber for part of the trip from the head end to the user, and have amplifiers at regular intervals to restore signal strength. In wireless, however, the signal strength drops from the base station toward the fringe of the cell and it is advantageous to employ a mixture of signal constellations.

1.5 Responsiveness to varying bandwidth demands

It is very important that the system be able to respond quickly to varying bandwidth requirements. Close tracking of the user requirements make is possible to run the system closer to capacity, without a need to allow a large margin for error. Also, in compressing out unused TDM channels, it is important to be able to restore them when needed. These goals are met by:

Using a relatively short scheduling interval of 1 millisecond.
Intercepting TDM signalling rather than depending on explicit requests
Piggybacking
Poll-me bit to indicate base station should poll user now
Adaptive polling lists with no polling of users already active

2 MAC Protocol Overview

The MAC protocol of this proposal includes several important features:

2.1 Request-grant allocation

Using a request/grant mechanism minimizes contention. At one-millisecond intervals, the base station broadcasts an allocation map which specifies which station is allowed to transmit when. This allocation map includes times for initial ranging (wide enough to accommodate station that do not yet know their distance from the base station) and times for making requests.

2.2 Polling

The base station stays informed of the needs of the user stations by periodic polling of several polling groups. A user is assigned to a particular group based on recent activity:

Inactive
Not recently active
Recently active
Currently active

Currently-active users are not polled unless they set the poll-me bit in the header when transmitting. Because they have the option of piggybacking new requests in transmission intervals already allocated to them, they generally do not need to be polled. Exceptions occur, for example, in the case of a station operating on a constant-rate basis, which is unable to substitute a request for current data traffic. In this case the station sets the poll-me bit to request a poll, so that it can request a bandwidth change or request an additional connection ID.

Recently-active users are polled more frequently than users that have not been recently active; and these more frequently than users that are registered but have not been active in a while. The polling intervals may be set up at the base station's option, but suggested intervals are every 20 milliseconds for recently-active users and every 100 milliseconds for not-recently-active users. Criteria for assigning the users to the groups are likewise the option of the base station.

2.2.1 Individual, group, and general polling

The polling process has a considerable amount of flexibility in how it may be done, to achieve the best balance between overhead and responsiveness. Polls may be addressed to individual stations or to multicast groups, or they may be broadcast to everyone. Standardization is not required with respect to the criteria for when users should be polled individually and when in groups or broadcast.

2.2.2 Contention resolution

In broadcast or multicast polling, or during registration, collisions may occur. In this case resolution is accomplished by the slotted-Aloha method, with the station waiting a randomized interval before retrying, with the interval doubling after each unsuccessful try.

2.3 Supporting various traffic types: convergence layers

The role of the MAC proper is to allocate transmission opportunities to the various users according to the service requirements of their applications, using the request/grant mechanism. This process is driven simply by byte count and quality-of-service considerations. However, other activities occur within the MAC layer, including registration, ranging, authentication, and encryption.

The formatting of data arriving at the MAC from higher layers, or other types of link layers, is the function of convergence sublayers. On an OSI Reference Model basis, these are part of the link layer, but their function is to format and pack the data rather than allocate bandwidth.

When transmissions are received, either at the base station or subscriber station, identification of the convergence layer to which the data should be handed up is implicit in the connection ID. (This is in contrast to LANs where the higher layer is identified in the Ethertype field or in the LLC SNAP header.) Therefore each connection ID is associated with a single convergence sublayer.

2.3.1 The virtual wire

The overall concept of the BWA network is that it looks to the subscriber like a wire or fiber: what goes in one end reappears at the other, without change. The only difference is that the shared nature of the point-to-multipoint system introduces delays in transmission.

Hence we adopt the mechanism of tunneling: data arriving at the subscriber station or base station are encapsulated and transmitted with the original headers and trailers intact. In the case of compressed TDM or ATM, framing or headers may be rearranged for transmission, but they are restored at the receiver.

Convergence layers supported include Ethernet, Frame Relay, IP, ATM, and video. The layer diagram is as follows:

Figure 1. Layer diagram showing convergence sublayers.

<table>
<tr><td>Eth, IP, PPP</td><td>ATM</td><td>(e.g. PBX)</td><td>Video network</td></tr>
<tr><td>LLC</td><td rowspan="2">ATM convergence</td><td rowspan="2">TDM convergence</td><td rowspan="2">MPEG convergence</td></tr>
<tr><td>Packet convergence</td></tr>
<tr><td colspan="4">MAC</td></tr>
<tr><td colspan="4">PHYSICAL</td></tr>
</table>

2.4 Data unit sizes

Data units carried may be of variable or fixed size. Optionally, a network may be configured to support fixed-size data units only; in this case the 802.16 headers are smaller since there is no need to convey the data unit size. Subscriber stations are informed of the network choice at registration time.

2.5 Duplexing

Three types of duplexing are supported: frequency-division duplexing (FDD), time-division duplexing (TDD) and half-duplex frequency division (HFDD). Choice of duplexing is a network option; both base stations and subscriber stations may support any one or more of the options.

The options are shown in Figure 2:

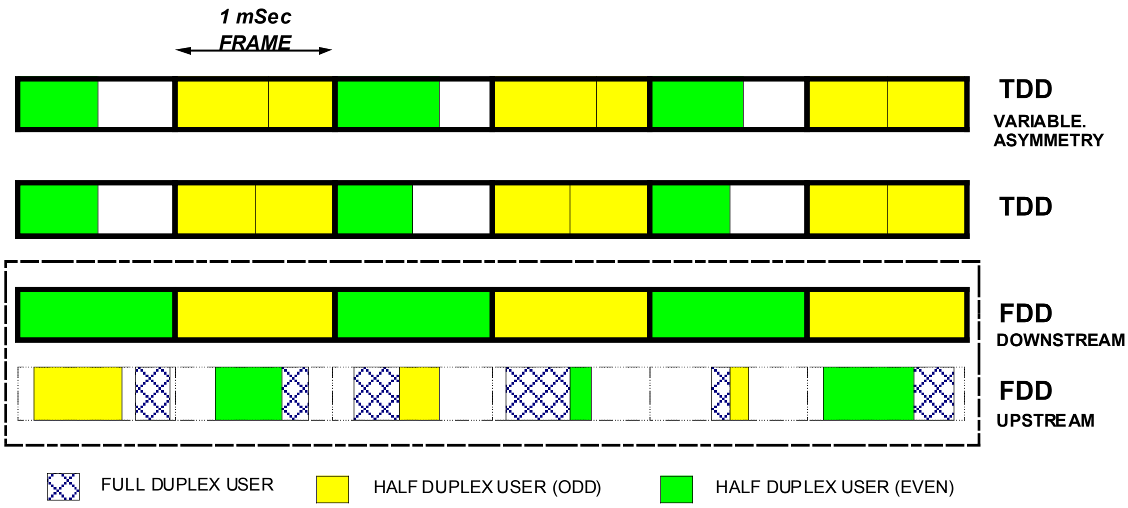


Figure 2. Duplexing options. Note that HFDD supports two sets of users 180 degrees out of phase.

The impact on the MAC of the duplexing choice is minor. In the FDD case, downstream and upstream are frame synchronized. A full duplex user may receive its downstream data and transmit its upstream data at the same instant, hence it may interact with the system on any given frame. A half-duplex user cannot transmit and receive at the same instant. Therefore if his downstream information is being received on even numbered frames then its upstream access occurs on odd numbered frames and vice versa. Both full duplex and half duplex users may co-exist in the same deployment and the MAC controller handles all users according to their limitations if any. In the TDD case, part of the frame is dedicated for downstream communications sub-frame and the remaining part for upstream communications sub-frame. The ratio between the downstream and upstream sub-frames may be fixed or variable to accommodate traffic asymmetry dynamics.

2.6 Compression

Compression is used to maximize the throughput of the air link. It is used in two areas: suppression of unused 64-kbps channels in the case of channelized T1 and E1, and in eliminating duplicate ATM headers. Additional compression may be employed at higher layers, for example video compression in MPEG, and IP header compression for voice over IP networks, but this is transparent to the 802.16 network.

2.6.1 Compressing TDM channels

While TDM is generally considered to be fixed-rate traffic, in fact the amount of usage on a T1, E1 or other TDM channel actually varies. In order to have sufficient capacity for peak periods, these circuits are usually provisioned with spare capacity most of the time; there is no need to carry inactive channels over the air.

Therefore the TDM channels are monitored in the TDM convergence process. When either side of the air link signals that one of the DS0 channels is going on-hook or off-hook, appropriate requests are made to change the bandwidth allocation for the connection. Since the response time is well within the circuit-setup allowance of the telephone network, call processing proceeds just as if the subscriber were hard-wired to the central office. (Of course, it is important not to overbook such traffic nor to book so many TDM connections that other data types are starved when TDM is fully busy.).

3 Addressing

Two types of addressing are used. The normal IEEE 802 48-bit format must be hard-wired into each base station and subscriber station; this address is used during the registration process.

3.1.1 Connection ID

In addition, a 16-bit temporary address known as a connection ID (CID) is used to associate data with a particular class of service and to designate data for a particular subscriber station and connection. Use of a small address field minimizes overhead.

It is expected that the physical layer of the BWA system will employ spatial reuse in the form of sector antennas; the full 360-degree spread can be divided into 6 or more independent systems, each with an independent instance of the MAC. In addition, frequency channelization may also be implemented, offering another degree of freedom with independent access control.

In each of these MACs, the CID space may be reused. While a single MAC may be limited to 2^{16} connections, because of the sector and frequency independence the total number can be many times that.

3.1.2 Basic connection ID

When a station registers, it initiates the ranging process with CID 0000. When the base station responds, it assigns the user station its basic CID. This CID is used from then on for exchanging control information. For other connections it must request additional CIDs. The CIDs assigned to a station are not numerically related to each other; they are simply 16-bit values that were available when the CIDs were assigned.

Requests for transmission opportunities are made by connection ID. This enables the base station to check the request for validity, QoS and for compliance to contractual limits. Since a user station may serve multiple customers within the building, each CID may have different service limits and billing arrangements. In such cases one or more different CIDs should be used for each subscriber, even for best-effort service.

3.1.3 Grants go to the station

The major exception to the use of the CID is in grants issued by the base station. While requests are made by CID, grants are issued to the user station only. This way the user station has the option of using the granted transmission time as originally intended, or to apply them to other CIDs. If new data for transmission has arrived at the user station with higher priority than the data for which the request was made, the user station can send that instead, making a new request to cover the original data. Similar to bandwidth allocation algorithms on the base station algorithms of grant utilization on the subscriber unit need not to be standardized.

4 Data Structures

The basic structure of the data transmitted in both directions is given by the 1-millisecond frame time. In TDD systems this represents the upstream plus downstream intervals; in both TDD and FDD this is the interval of transmitting the allocation map.

4.1 Downstream subframe

The base station sends out a transmission stream that starts out with the PHY frame marker, then frame control information in the most robust modulation (QAM-4). This is followed by data for the subscriber stations, organized by modulation. Data for the most distant users is sent first in QAM-4, then a block of data for QAM-16 users, and finally data in QAM-64. This is shown in Figure 4.

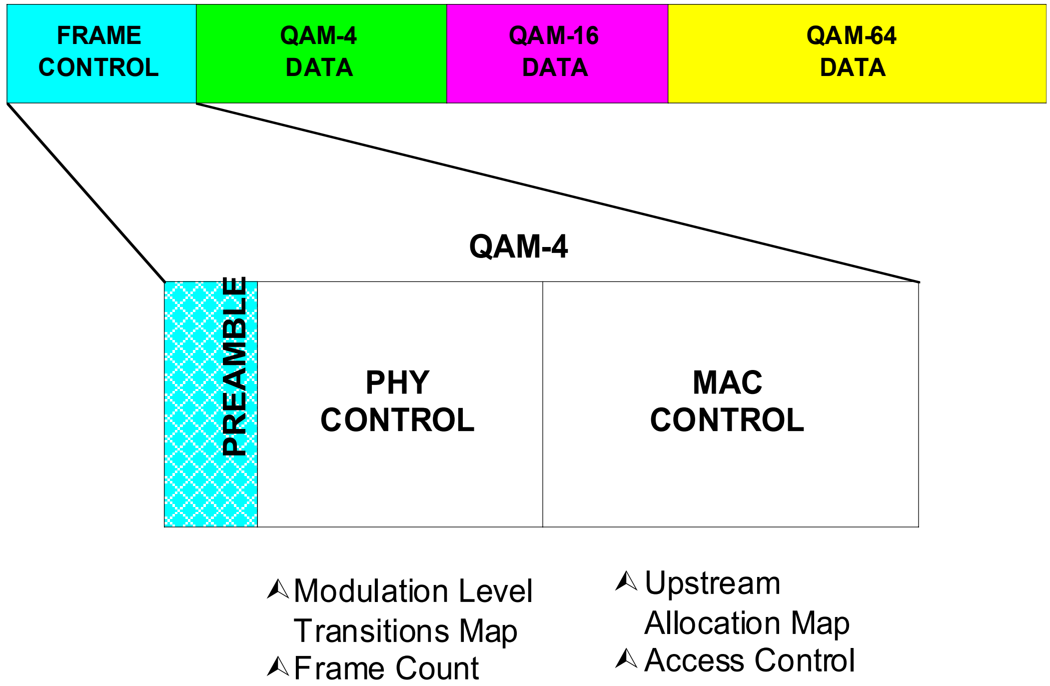


Figure 4. Downstream subframe structure.

4.2 Upstream Subframe

In the upstream direction, transmissions are likewise organized by modulation type, progressing from QAM-4 to QAM-64 (as shown in figure 5). Due to physical layer overheads of guard time and data preamble, each station must transmit all of its data for this frame (not exceeding its grant) in the same burst, regardless of priority.

Note that while scheduling algorithms per se need not be standardized, the grouping together of all transmissions by a given user station must be a result of the process.

In addition to the user transmissions, the upstream subframe includes periods for registration and for bandwidth requests. While registration is a general-contention process, the bandwidth request interval may be subdivided by the base station into multiple individual or group request periods. In all of these cases the registration or request transmission by the user must take place at a randomized time within the interval specified for the purpose. The structure is shown in Figure 5.

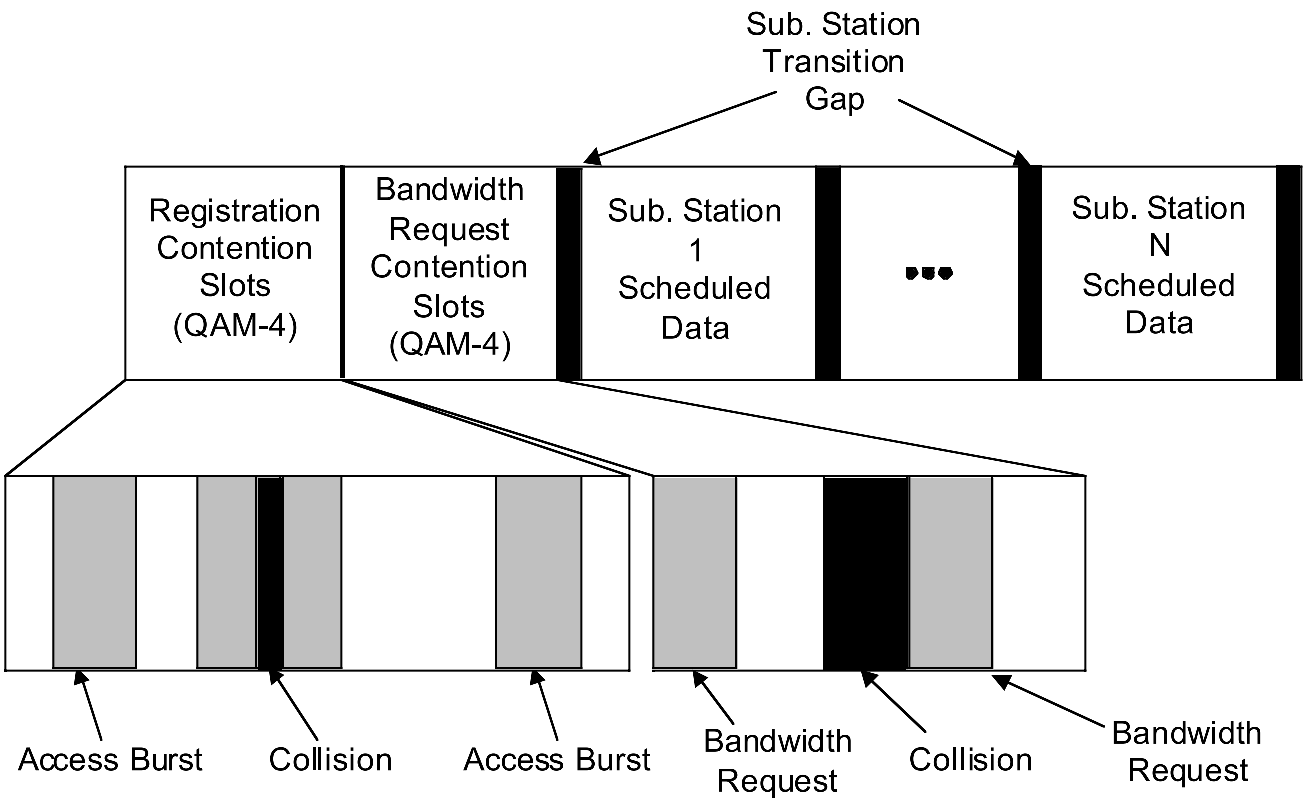


Figure 5. Upstream subframe structure

When collisions occur, the data are corrupted which must be ignored by the receiver. The result is that no response will be returned to the sender, for example no grant in the forthcoming allocation map.

4.3 Packet formats

The data format is the same in the upstream and downstream directions. The only exception to this rule is the power control field, used in the downstream direction for the base station to request an adjustment of the user's power level. In the upstream direction, this becomes the poll-me field.

Two options exist for the data units: variable and fixed length. The fixed-length option does not need a length field; otherwise the formats are the same except for the Backhaul Reserved (BR) field. The Backhaul Reserved field is used to carry information relevant to the backhaul and which may be ignored by the MAC. For example if the backhaul uses ATM, the BR field would carry the PTI information across the air interface.

While the fixed-length data units may be ATM cells, they may also be any length that does not change, up to a limit of 2048 bytes. Note that ATM header compression cannot be used with the fixed-length option, since an unpredictable number of ATM headers may be suppressed.

4.3.1 Variable-length packets

The format for variable-length data units is shown in Figure 6. Encoding of parameter values for the various fields will be provided later. Fixed-length data units may be carried if desired, with the length specified explicitly.

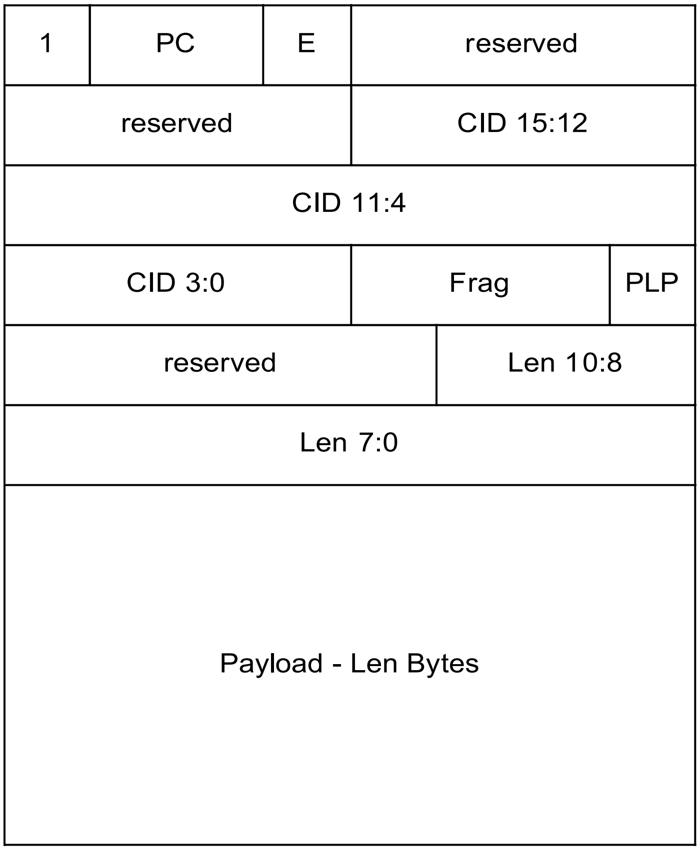
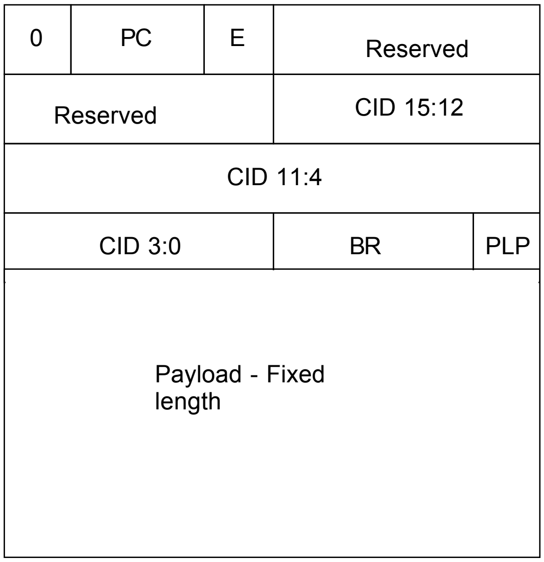


Figure 6. Format for variable-length data units.

4.3.2 Fixed-length packets



PC	power control (downstream)/ poll me (upstream)
E	encryption indicator
CID	connection ID
BR	Backhaul Reserved
PLP	packet loss priority

Figure 7. Format for fixed-length data units.

4.4 Packet fragmentation

User packets may be interrupted by the end of the frame or the end of the allocated transmission opportunity. In this case, the length field applies to each fragment individually; this value is always expressed in bytes. Indicators in the header specify whether the fragment is the beginning, continuation, or end fragment, or whether the data unit is a single segment.

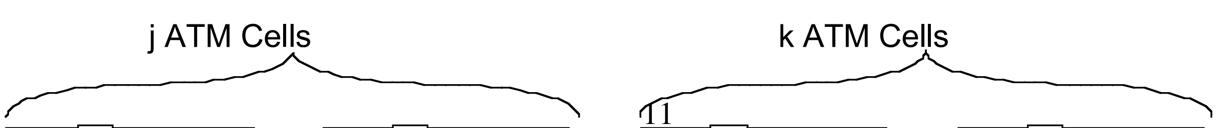
Beginning	01
Continuation	00
End	10
Unfragmented	11

When ATM cells are compressed, the result is independent of the fragmentation; fragmentation can occur in the middle of a cell. The ATM convergence process must be prepared to receive fragmented and compressed cells.

4.5 Optional Feature: Compression of ATM Headers

In return for its benefits in flexibility and provision of varied qualities of service, ATM exacts a penalty in bandwidth of approximately 10% for the cell headers. However, in practice a burst of cells with identical headers enters ATM networks, due to the segmentation of a single data packet into a multiplicity of smaller ATM cells. Such redundancy makes the ATM headers a candidate for compression.

Compression is accomplished by sending the first header in a series of identical headers, along with a run-length count. If ATM header compression is performed it shall be done within frame boundaries (e.g. 1 millisecond). The most-significant bit of the 8-bit count field is used to indicate that the last cell of this compression group is the last cell of the packet, i.e., that the AAL 5 last-cell bit is set. Hence there is no need to send this header as a separate compression group even though it differs (in this bit) from the previous header. Header compression is shown schematically in Figure 8.



HEC

Fig. 8. Compression of identical ATM headers into a single header plus run length.

4.6 TDM Support

Efficient support of TDM traffic is maintained by transmitting only the active DS0 channels, in cases where the trunk is channelized. If the trunk is not channelized, for example if it is used for data traffic, then no compression is possible.

Interpretation of the signaling is the responsibility of the user station. This is desirable because there are many variants of the signaling protocols in use. Rather than require the base station to be conversant with all of them, each user station is responsible for monitoring the particular variant in use by the attached equipment on the user's site. The user station modifies its bandwidth requests as it detects channels going off-hook or on-hook. Grants of transmission opportunities are on a continuing basis for TDM: they are valid until changed without a need for new requests, though the grants still appear in each allocation map.

For example - a T1, in a one-millisecond packetization interval up to 192 bytes of data are carried plus 3 bytes of Time Slot Map. The Time Slot Map indicates whether a channel is active or suppressed.

Formats for TDM packets are given in Figure 9.

Compressed T1 with channel-associated signaling

Header	Time slot MAP 3 bytes for T1	8 TDM sub frames

Compressed E1

Header	Time slot MAP 4 bytes for E1	8 TDM sub frames

TDM Subframe

1st active DS0	2nd active DS0	3rd active DS0	…	Last active DS0
←--- 125 us -------------------------------------→				

Fig. 9. Data formats for various types of compressed TDM traffic.

Both ends of the air link monitor the state of the Robbed Bits Signalling (RBS). Changes in the RBS state are sent using and out of band control channel.

Since the BWA frame structure imposes a 1-millisecond granularity on the traffic from any individual user station, the base station must be prepared to provide a de-jitter buffer that will permit receiving a millisecond's worth of TDM data from each station and feed it out into the backbone network with conventional timing (normally one byte per DS0 at 125-microsecond intervals).

5 Scheduling

5.1 Upstream

Upstream traffic, from the user station to the base station, is connection-oriented in that it operates with a saved context, that of the request/grant mechanism. This is true for all traffic including nominally connectionless service like IP packets. In addition, other connection-oriented traffic such as ATM, may be tunneled through this link; it may terminate well beyond either end of the wireless link.

Because of the physical-layer overhead, it is necessary to schedule all transmissions from a given user station in the same burst. The user station is allocated a transmission period sufficiently long to accommodate all requests (explicit and continuing) received during an interval of one frame time, unless the sum of all requests exceeds the available transmission time. In this case, a user station may be given less than it has requested. In some cases, if a stations requests are of low priority, it may not be possible to allocate it any bandwidth at all in the current frame. Unfilled and partially filled requests must be carried over for the next frame.

While the base station performs allocation on the basis of requests received, it transmits in the allocation map a total allocation for each user station. This enables a user station to transmit new data of higher priority than that for which allocation was requested. A new request can then be made to provide for the original data.

While precise details of the scheduling mechanism is not needed for interoperability, some general guidelines may be in order. The base station should organize incoming requests by both priority and user station. Building the allocation map can then proceed with accommodating user requests in priority order until the available transmission time is exhausted. The order should be:

Constant rate
Variable rate real-time
Variable rate non-real-time
Available rate (in the ATM sense)
All other traffic (generally best-effort data)

Any remaining unallocated requests will generally be at low priority.

In sending allocations, the base station must group together user allocations by modulation type: all QAM-4 together, etc. This avoids the need to reconfigure its receiver for every incoming user burst. Although appropriate equalization must be done for each user station, the process is faster than attempting to determine on the fly what type of modulation is being received.

Registration and request intervals are also incorporated in the allocation map. Polling is accomplishing by specifying the request intervals, which can be addressed to individuals stations, groups, or broadcast to all stations. Any or all of these can appear in the same map, as the base station optimizes use of the upstream bandwidth. An example of the allocation map is shown in Figure 10.

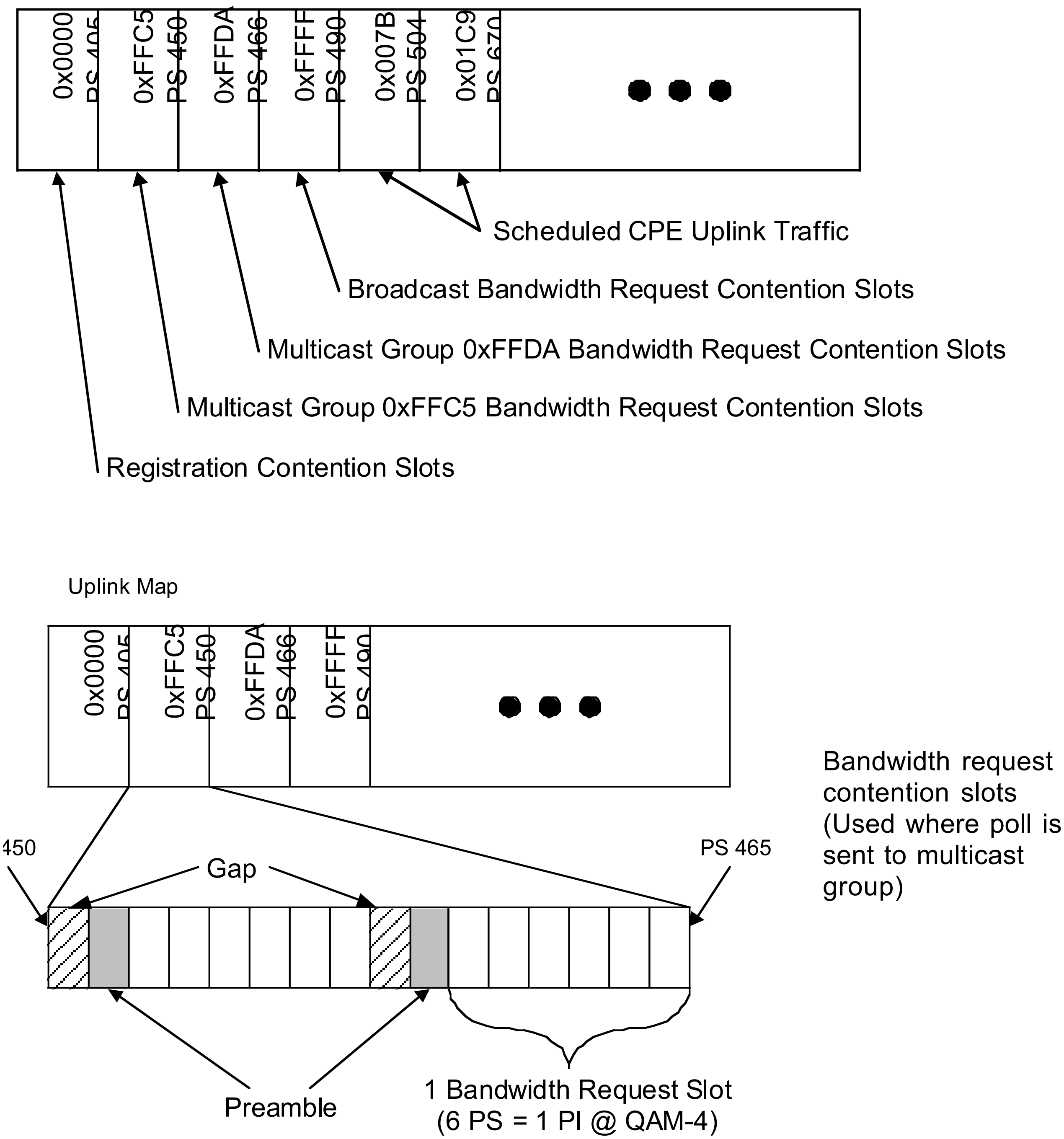


Fig. 10. Format of bandwidth allocation map.

5.2 Downstream

Downstream scheduling is similar to upstream scheduling insofar as the base station is concerned. It receives data for the forthcoming frame and organizes it by user station in order to check compliance with user service contracts, and also to group the data by modulation type.

Since there is only one downstream transmitter, there are no issues of changing from one transmitter to another, as there are upstream. Without PHY overhead, it is equally efficient to transmit one time or multiple times to the same user station. The base station may do this at its option, for example to minimize jitter for TDM service.

Data for a given user station are recognized and copied based on CID rather than the 48-bit hardware address.

5.3 Polling groups

Other than constant-rate traffic with continuing allocations, it is assumed that user traffic varies in rate. The user requirements are determined by polling: providing allocations in which the user may make requests. In the case of an individual poll, the user station must reply, even if the amount requested is zero. This acts as a keep alive and helps keep the equalization parameters for each user up-to-date and minimizes equalization time when there is data to be sent.

As indicated in the MAC overview, users can be assigned to several polling groups:

Inactive
Active, but not recently
Recently active
Currently active

Polling frequency should vary, with inactive users polled least often, or polled on a group basis. Currently-active users need not be polled since they can either piggyback a new request in an existing allocation, or they can use the poll-me bit to request a poll when they need to make a change. Polling strategies are not subject to standardization.

Polling overhead is minimal, as the following calculation shows. The result is upstream overhead of 0.6% and downstream overhead of 0.05%.

Assume that there are 100 user stations in this sector. Assume 12.5-MHz FDD channels with 10 Msps (megasymbols per second) or 25-MHz for TDD with 20 Msps, providing 30, 60, or 90 Mbps for QAM-4, -16, -64 (payload, not counting FEC), and that the average is 60 Mbps in each direction.

Assume that 50% of these users are inactive, polled occasionally on a group basis. Overhead is minimal.

Assume that 25% are not recently active; their polling interval is 100 milliseconds.
The upstream poll response is 308 bits: 208 bit PHY granularity + 100 bits guard & preamble time (based on QAM-4). Then the overhead is 25x10x308 bits = 77 kbps.
Downstream, polling requires 24 bits per poll, resulting in overhead of 25x10x24 = 6 kbps

Further assume that 15% are recently active; poll them every 20 ms.
Upstream overhead is 15x50x308 = 230 kbps; total, both groups = 307 kbps.
Downstream overhead is 15x50x24=16 kbps; total, both groups = 22 kbps.

The remaining 10% of users are already active and don't need to be polled.

The total overhead is then 384 kbps upstream, or 0.6%, and 28 kbps downstream, or 0.05%.

6 Registration

The registration process is similar to that used for cable modems. A newly powered-up subscriber station must find the correct frequency, if several are in use, and proceed to synchronize itself with the base station. This process must start at the physical level and work itself up through the protocol stack: no layer can initialize itself properly unless the lower-layer services that it depends on are available.

The sequence of the process is as follows:

Set to most recent frequency channel and sync on modulation
- Sync on framing
- Sync on downstream MAC messages
- Get parameters for upstream transmission
- Wait for allocation map to locate initial ranging interval
- Do distance and power ranging with connection ID = 0
- Get basic connection ID and time of day
- Get IP address with DHCP for use with SNMP
- Authenticate station using public key encryption
- Exchange session keys

More details will be provided in the future.

7 Security

The wireless environment is subject to easy unauthorized physical access, more so than any other medium; in this case there is no way to detect that signal energy has been diverted to unauthorized ends. Hence security must be a mandatory feature of the wireless access network.

The purpose of 802.16 security measures is to protect the air link. Additional measures may be provided over other links in the overall path, or on an application-to-application basis, where the level of security can be tailored to the application.

Several security procedures are needed to ensure network integrity and privacy of user communication:

Authentication of user
- Privacy -- preventing eavesdropping
- Integrity -- preventing alteration of data

Registration procedures are used to authenticate a newly power-up user station. These are based on a public-key process and involve the hardware ID of the user station. Session keys are replaced periodically.

User data is encrypted using DES or triple-DES, with headers in the clear to route data to the correct user. Control information is generally not encrypted. An encryption indicator bit is set in header for encrypted-payload packets, to facilitate hardware processing. While DES is not necessarily state of the art in encryption, export is relatively straightforward, and hardware to perform the encryption and decryption is readily available.

8 MAC Messages

Messages used at the MAC level include:

Message	Direction	Purpose	Code
PHY Control	BS → SS	Physical Layer Control	0x00
MAC Control	BS → SS	MAC Control	0x01
Registration	SS → BS	Registration	0x10
Registration Results	BS → SS	Registration	0x11
Ranging	SS → BS	Registration	0x12
Ranging Results	BS → SS	Registration	0x13
Re-register	BS → SS	Registration	0x14
Registration Collision	BS → SS	Registration	0x15
Change Modulation	BS → SS	Physical Layer Maint	0x21
Mod. Change Ack	SS → BS	Physical Layer Maint	0x22
Tx Advance Change	BS → SS	Physical Layer Maint	0x23
Tx Advance Ack	SS → BS	Physical Layer Maint	0x24
Power Adjustment	BS → SS	Physical Layer Maint	0x25
Power Adjustment Ack	SS → BS	Physical Layer Maint	0x26
Bandwidth Request	SS → BS	Connection Maint	0xB0
Multicast Assignment	BS → SS	Connection Maint	0xC0
Multicast Assign. Ack	SS → BS	Connection Maint	0xC1
Key Sequence	BS → SS	Security	0x40
Key Sequence Ack	SS → BS	Security	0x41
Channel Change	BS → SS	Load Leveling	0x50
Channel Change Ack	SS → BS	Load Leveling	0x51